



Preliminary Draft
Not For Quotation

PAKISTAN INSTITUTE OF DEVELOPMENT ECONOMICS

Seminar Paper No. 26

October 1977

AN ECONOMIC ANALYSIS OF PERSONAL EARNINGS
IN RAWALPINDI CITY

Nadeem Ul Haque

AN ECONOMIC ANALYSIS OF PERSONAL EARNINGS IN RAWALPINDI CITY

by

Nadeem Ul Haque^{*}

INTRODUCTION

The intent of this paper is to delineate the determinants of the distribution of income in Rawalpindi city. The basic hypothesis to be tested is that for each individual, personal earnings are a function of his socio-economic characteristics ie. age, sex, education etc. If the hypothesis bears out, the disparity of incomes within the city would then prove to be a consequence of the variation in these individual characteristics. It would then be of interest to quantify the effect of each of these determinants on the eventual distribution of income.

The paper has been divided into six sections. The first sets out the methodology and the theoretical framework for the analysis. The next section looks at some previous research which is of relevance to the topic. In section 3 the data set is described and some reservations about it discussed. The results are presented in sections 4 and 5. In the conclusion some econometric problems and the main results of the analysis are discussed.

^{*}The author is a Research Economist at the Pakistan Institute of Development Economics. Much gratitude is owed to K. Hamdani and Dennis Detray for the considerable help and advice that they were always willing to give.

Theoretical Framework

The basic framework for the following analysis is provided by the theory of human capital. The corner stone of this theory is that human beings invest in themselves in a variety of ways i.e. incur present costs for future benefits. Education, on-the-job-training, work experience are all important manifestations of this phenomenon of self-investment. The acquisition of human capital raises an individuals productivity, and because employers pay in proportion to productivity, the individuals earnings. Each person invests in human capital to the point where the marginal return equals marginal cost.¹

In this study education is used as a direct human capital variable while age is taken as a proxy for the human capital variable work-experience.² The hypothesis therefore is that personal earnings are a positive function of age and education. This relationship is expected to explain much of the observed variation in earnings.

However the observed relation between earnings on the one hand and education and age on the other, may just be a statistical illusion. For example if only the children of the rich can afford to go to school, then the well paid jobs which they might get after finishing school may not be a result of the amount of human capital acquired, but of their family connections. Similarly if only the

¹Becker "Human Capital and the Personal Distribution of Income.

²In the human capital specification education should be a continuous variable measured as the number of years of formal schooling. The Rawalpindi Survey however picked up only the educational level attained.

very bright get educated then higher earnings are a consequence of innate intelligence and not education alone. Again structural defects in the market explain the disparity in incomes. Women may be discriminated against, irrespective of age and education. The prevalence of the dual markets might hinder the human capital process at work. In this case these human capital variables will be of relevance only in the formal sector, and not in the informal.

There are therefore likely to be a large number of variables which determine the eventual distribution of income. The use of multiple regression analysis will allow us to separate the effects of a number of independent variables on a particular dependent variable. This technique will enable us to explain the variation in personal earnings in terms of variation in individual characteristics such as education, age, sex, occupation etc. Our model therefore is the classical least squares equation with both dummy and continuous regressions.

$$y_i = \alpha + \sum_{j=1}^n \beta_j X_{ji} + u_i$$

where j stands for a variable and i stands for an individual, y_i is the natural logarithm or level of earnings or the wage rate, $X_{1i} \dots X_{ni}$ are n observable characteristics (continuous or dummy) used to explain y_j , and u_i are the random unobserved disturbances with zero mean and constant variance.

The regression model will identify and rank for us the determinants of the distribution of income. The regression package will at the same time provide us with the means and variances of the variables used in the

equation. Considerable useful information about the sample or any break down on the sample that we may wish to consider, can be obtained from these summary statistics. Of particular interest would be the log variance of income a much used measure of income inequality³.

Being a relative measure of inequality, the log-variance will be useful to us mainly when making sub-sectoral comparisons, i.e. when we consider various sub-samples.

PREVIOUS RESEARCH

Income distribution studies in economic literature tend to be one of three distinct types. The first is concerned with the statistics of the observed distribution of income. Statistical distributions which best fit the observed distribution of income or the distribution above a certain level, are examined. Underlying stochastic processes which could generate such distribution are also studied⁴. For the second type the inequality measures like the Gini coefficient are calculated. Comparisons of these measures amongst nations, amongst regions within nations or amongst various grouping within a population are then made in an attempt to understand inequality⁵.

The third type of study is relatively new and unlike the other two is derived from economic theory. This is the theory of human capital, that has already been discussed. As noted earlier, here an individual's

³See 'Sen' [9] about the merits and de-merits of the measure. Basically the log variance is a relative measure of income inequality. The measure is free of any change of units. It is sensitive to income transfers at the lower end of the scale and therefore useful to us as we are probably dealing with the middle portion of the income distributions.

⁴See for example D.G. Chanpernowne "A model of income distribution" Economic Journal June 1953, Maadelbovt, "The Pareto Levy Law and the distribution of income", International Economic Review. May 1960.

⁵James Morgan "The anatomy of income distribution", Review of Economics and Statistics. P.D. Ojha and V.V. Bhatt "Pattern of income distribution in an underdeveloped economy. A case study of India". American Economic Review September 1964.

earnings are a positive function of the amount of human capital he possesses⁶. The distribution of income is therefore a consequence of individual supply and demand for human capital.

In Pakistan, research on the distribution of income has been mainly of the second type, i.e. the calculation of inequality measures. In her study on the "Measurement of Inequality in Urban Incomes in Pakistan", Khadija Maq calculates the Gini coefficients for the years 1948-49 to 1960-61 and estimates the trend over time of inequality⁷. For her data, she uses the estimates of personal income based on the "All Pakistan Income Tax Returns" published by the CSO. Her analysis is therefore restricted to the very high income groups, (over Rs. 35.00 per month) and some 0.01% of the population. For this income group, her results show that (a) income is more inequally distributed in Pakistan than most developed and developing countries, (b) income is shelved in favour of the rich but the trend is towards the reduction of disparities within the high income bracket⁸. Her analysis however does not include any estimate of leakages due to misreporting of taxes. Also over time the proportion of fringe benefits in personal income may have been increasing.⁹ An inclusion

⁶Gary S. Becker "Human Capital and the Personal Distribution of Income. J. Mincer. "The distribution of labour incomes. survey". Journal of Economic Literature March 01, 1970.

⁷She calculates both the Gini coefficient and the Puarcto coefficient for all the years between 1948-49 and 1960-61. Both these coefficients reveal a negative trend. The Gini coefficient falls from 0.61 in 1948-49 to 0.45 in 1960-61.

⁸It would be fair to point out that she qualifies these results rather heavily. She sets out clearly the limitations of her data. On her results the relative position of new entrants and lower income groups in the tax-paying population is improving. She also notes that the tax-paying population has been receiving an increasing share of the national income while paying a decreasing percentage of it in taxes.

⁹The author is concious of these.

of either of these may give an increasing trend for inequality¹⁰.

For our purposes, a more interesting study and one to which frequent reference will be made is Blaug's study of earnings in Thailand. Blaug had a very large sample, about 9000 observations¹¹ " A large amount of detailed information was collected for each individual. For example questions was asked on all the jobs held by an individual; fringe benefit in cash or kind; social background of respondents; the type of school attended occupations, hours of work; sector of employment and even income from property and self-employment.

From those individual characteristics Blaug derived 69 new variables. Using a specification similar to the one used in this paper, and stepwise regression procedures he sifted out the effects of these variables on income the dependent variable. The results of his analysis were:-

- (i) An almost linear age-earnings profile. There was a very shallow concavity in the age earning profile which could not be picked up by the normal human capital specification using age and the square of age.
- (ii) Education was an important determinant of the distribution of income. The hypothesis of a positive relationship between the two was accepted.

¹⁰ The only other study on personal income distribution for Pakistan was Ashbjorn Bergen's "Personal Income Distribution and Personal Savings in Pakistan 1963-64". He measures the Gini coefficients for East Pakistan, West Pakistan, rural areas urban areas etc. Using the CSO's Quarterly survey data. He also calculates saving ratios. His main conclusion however is that the data base is too weak for the analysis he was trying to do - Most other studies in Pakistan deal with national income and its distribution.

¹¹ He used the household economic survey of 4,600 observations. As college graduates were under-represented, another survey was conducted. 200 random observations and 3000 men and women were "interviewed purposively to reach predetermined quotas defined in terms of age sex and education.

- (iii) The two human capital variables, education and age combined explain most of the variation in income¹².
- (iv) Amongst the education variables, the higher levels of education were the ones that contributed the most towards explaining the variation in income.
- (v) Family background, employment status and occupations were not insignificant when explaining income distribution.

Another study of interest is Sudbir Anand's "Size Distribution of Income in Malaya". Anand too had a large sample, 6000 observations. He was not concerned however with estimating the effects on income, of variables other than the human capital. In fact he was mainly testing the applicability of the human capital model in Malaya. The intention was to see how well the model explained the disparity in incomes. He used what may be termed a "pure" human capital specification.

$$\text{Log } Y = B_0 + B_1 S + B_2 T + B_3 T^2$$

Where

Y = Income

S = Years of schooling

T = Years of labour force experience and was measured as
 $T = \text{Age} - S - 5$

The equation was run for the whole sample and for various subsamples, were selected for the occupations, the sexes, various age groups, social group and educational levels. The results indicated that the human capital model explained a large part of the variance in incomes in Malaya. The basic hypothesis of increasing returns to education and age and the concavity of the age-earnings profile over all supported by the analysis.

¹² Blaug includes sex in his "basic" variables and it is the three of these prove to be most significant.

THE DATA

The data used in the analysis is the Rawalpindi Socio-Economic survey conducted by the PIDE in 1975. As a detailed description of the survey the sampling design and the possible sampling and non sampling error is found in Hamdani [7], this discussion will only concentrate on the points of interest to us. Briefly basic socio economic information was collected from a thousand households in Rawalpindi. As Hamdani notes, "a tight budget necessitated the small sample size and the simple sampling design". He concludes however, that "the sampled households are representative of Rawalpindi".

A brief summary of the data is as follows. The labour forces consists of 1641 individuals of whom 1541 are males. A little less than half of the labour force 49.24 % are regular employees while 33.4 % are self-employed. Surprisingly only 1.4 % reported as being casual employers and 5.3 % as un-employed. Under-employment however was substantial : 23.8 % would like to work more hours¹³. A minority of the labour force are apprentices, 3.2 % and unpaid family helpers., 7.5 %.

The use of this data for an income distribution analysis however, must be made, bearing in mind certain reservations. A brief description of the city will bring out the first of these. Rawalpindi is the fifth largest city in Pakistan with an estimated population of 6,73,000 individuals in 1975¹⁴. An important regional metropolis with primarily administrative functions, the city was the country's interim capital in the early sixties. The development of the country's new capital Islamabad on the outskirts of Rawalpindi has aided the development and expansion of its wholesale trade and construction

¹³This is however, a very unsatisfactory measure of under-employment.

¹⁴The figure was arrived at by projecting the 1972 census estimate at a 3.2 % annual growth rate.

activities. Manufacturing activity is virtually non-existent in the city. Essentially the economic activities of the people in the city would be trade, construction and administration (i.e. government employees)¹⁵. The absence of manufacturing activity and the city's close proximity to the national capital make it somewhat unrepresentative of the other urban areas of Pakistan. To that extent our results would be particular to Rawalpindi.

A bias stems from the sampling design. The sampling frame consisted only of structured and un-structured dwellings. This probably resulted in the data yielding an income distribution truncated at the lower tail i.e. not enough representation of the very poor in the sample. If housing were to be regarded as a normal good, with a positive income effect, then it may be assumed that the very poor live in unstructured or kutcha dwellings.¹⁶ Our sample in that case underestimates the inequality or the variation in incomes.¹⁷ The distinction between structured and unstructured dwellings however fades if we accept what is a fact in underdeveloped countries that the structured dwellings of the poor tend to be over-populated slums. In fact the mean income for our sample is Rs. 376 which is really very low especially when we take into account that there are on an average 3.29 people dependent on an earner. This makes the per capita income in the

¹⁵To this must added a sampling bias. Rawalpindi has had since the pre-independence days one of the largest army cantonments. The military personnel were however deliberately excluded from the sample because it is illegal to gather any information on them.

¹⁶It could also be that living in the unstructured dwellings may be recent migrants earning about the same as the others but in transit to move into the structured houses.

¹⁷As mentioned earlier the mean income for the sample is lower than the per capita income. The sample probably did not pick up enough of the rich. In either case there is sample reason to assume that the inequality has been underestimated.

Successively high levels of education command increasingly higher levels of income, i.e. difference between the coefficients of higher and secondary levels of education is greater than that between the coefficient for secondary and primary education 0.394 and 0.260 respectively. Amongst the education categories primary education has the smallest coefficient in all the regressions. The coefficients had however become small, insignificant and negative, when regression was tried for all earners, including the unemployed apprentices and the family helpers²⁶. This could be due to there being more of the primary educated amongst the unemployed and the other excluded classes than those from the other education categories.

²⁶Regression 5 was tried because this was the only specification permitted by the data. Remember we have no occupation and employment status record for these excluded classes.

The results are:

$$\begin{aligned} \text{Log } y &: 0.418 + 0.221 X_1 - 0.002 X_2^2 - 1.344 X_{12} \\ &\quad (0.017) \quad (0.0002) \quad (0.176) \\ &- 0.619 X_{13} + 0.0006 X_{14} + 0.666 X_{15} \\ &\quad (0.125) \quad (0.109) \quad (0.152) \\ &+ 1.001 X_{16} + 0.195 X_{18} \\ &\quad (0.206) \quad (0.104) \\ n &= 1641; R^2 = 0.337, F = 103.57 \\ \text{mean } y &= 376.436 \text{ std. dev. } \log y = 2.185 \end{aligned}$$

These results are reliable to the extent that the probability of being unemployed given by the sample reflects the actual probability. The coefficients of X_{15} , X_{16} and X_{18} now all rise in value. This is probably because inclusion of the low income groups has now decreased the value of intercept term. The mean income is also now lower than before. The returns to secondary and higher education therefore are all earning lesser than before probably because of their larger representation in these low income groups.

Of the education coefficients technical education is the only category that presents a surprising result. Being a human capital variable, the attainment of technical education was also expected to raise earnings. In all our regressions however, the coefficient for this category is not significantly different from zero. This gives us the rather startling result that those with technical education earn no more than those without it. The only meaningful definition of this variable that the data would allow was more than six months of on the job training or apprenticeship. Formal technical education is available in very limited supply in Pakistan. Most skills are gained through a very long period of apprenticeship. Not surprisingly the drop-out rate is very high. The regression coefficient may thus be indicating this phenomenon.

From the employment status variables, the excluded category is the regular employees. Of the included categories the self-employed with employees have the largest coefficient and the most significant. Individuals in this category (some sixty of them) earn about 58.5% more than regular employees. However for these "richer" self-employed it is probably reasonable to expect that part of their earnings is a return on capital.

A somewhat surprising result is that for the self-employed without employees. These individuals as it turns out earn just about the same as the regular employees. The casual workers however according to expectation earn about 20% less than the regular employees. Roughly speaking the two categories, the self-employed and the casuals constitute what is known as the informal sector. Theory suggests that individuals

within these categories desire to get into the formal sector but market imperfections prevent them from doing so. For the self-employed higher earnings it seems is not one of the incentives for wanting regular employment. For the casuals however it must be noted that only those who admitted to having worked for at least the past week were admitted into the category and their work - characteristics of the past week were recorded. No employment history for the past year or any long period was recorded. It would however not be unreasonable to assume that these workers would face periods of unemployment during a year. Average earning over the year for the casual should therefore be even lesser than those indicated by the regressions. Both the casuals and the self-employed however put in much longer hours than regular employees. The opportunity of any secondary employment is therefore severely limited for the former. Also these individuals are affected by the vagaries of the market whereas the regularly employed are not.

For the migrant, the results are both the most surprising and the most interesting. The expectation in this case was that migrants in their early days of migration would probably earn less than the residents but gradually upon acquiring location specific human capital they would earn as much if not more than the locals. In the long run migrants are expected to earn more than the residents for migration it must be remembered is a self-selection process whereby only the more dynamic and the more capable are thrown up. Our results however show the recent migrants to be earning more than those who have settled in the city: the new arrivals about 20.1% more than the inhabitants of the city, the relatively settled 23.1% more while the settled ones only 12% more.

An explanation of these coefficients can be the increasing economic importance of Rawalpindi in recent years because of the development of the new capital on its outskirts. The city in the recent past has been offering better opportunities and thus providing incentives for better qualified people to migrate to it. Therefore existing earnings could be a reflection of differing market conditions at the time of migration.

Another surprising fact about migrants is revealed by the following cross-tabulation.

	<u>Regular Employeeed</u>	<u>Self Employed with employeeed</u>
Migration 1-3 years	84.5 %	15.5 %
Migration 4-6 years	77.6 %	22.4 %
Migration 7-15 "	56.2 %	38.8 %

The dual economy suggests that most of the migrants especially the recent ones would be found in the informal sector which by definition corresponds to our self-employed with employees. The above results however contradicts this hypothesis. A large majority of the recent migrants it seems are able to find regular employment. This could be indicative of any number of things; for example are these immigrants more productive than the others to be able to get regular jobs so quickly? Is it a result of the government's regional policy that these people get better jobs? Is the earning pattern indicative of this distribution amongst the sectors? etc.

Returning now to the results in table 2 contd. note that we are using the unconstrained wage rate specification in our regressions. Economic theory suggests that the wage rate be used as the dependent

variable for it is this variable on which the individual labour supply decisions are based²⁷. Although we do not use the wage rate explicitly as a dependent variable, however controlling for hours worked in essence means that we are calculating a wage rate equation. An additional benefit is that we get a coefficient for hours worked²⁸. A comparison of regression 1 to 6 illustrates the above point. Both regressions use the same number of variables and the same functional form the only difference being in the dependent variable. It is evident that differences where they exist are not alarming.

The coefficient of hours worked the results indicate is 0.190. An elasticity coefficient, this should be interpreted as a 10% increase in the number of hours worked producing a 1.90% increase in income²⁹. Decreasing returns to extra hours worked are indicated. Had the coefficient been equal to one we would have had an exact wage rate equation, and proportional returns to working additional hours. The low value of the coefficient is probably an indication of the costs of leisure being very high. People are probably willing to work at differing and low

²⁷ Each individual works upto the point where his marginal rate of substitution of leisure for money is equal to the wage rate.

²⁸ Author reason for using this specification is that we are not sure of our hours worked variable. In fact there is reason to believe that there may have been some misreporting.

²⁹ Income is a concave function of hours worked.

wage rates in order to maximise total earnings³⁰.

Amongst the occupations the highest paid are the administrative and managerial workers. These are followed by the professional and technical workers, the sales workers, the agricultural workers³¹, the clerical and related workers and the service workers. The last mentioned are the only ones with a negative coefficient. The excluded were the production workers. A surprising result is that clerical, service and production workers should earn as much as the dairy workers. In regression 6 however all other occupations become significant except for agricultural workers, service workers it turns out earn definitely less than production workers³².

³⁰ Another reason for this low value for the log of hours worked coefficient is that the regression includes only those who are gainfully employed. The observations are all centred around the mean hours worked. We may therefore be capturing only a small segment of the hours worked-income curve. Introduction of the part time workers and the unemployed may raise the value nearer to one. The following regression rather imperfectly illustrates this point.

$$\begin{aligned} \log y = & 2.688 + 0.174 X_1 - 0.002 X_1^2 - 1.084 X_{12} \\ & (0.015) \quad (0.0002) \quad (0.122) \\ & - 0.450 X_{13} + 0.263 X_{14} + 0.850 X_{15} + 1.33 X_{16} \\ & (0.108) \quad (0.095) \quad (0.101) \quad (0.178) \\ & + 0.158 X_{18} + 0.769 \log X_{33} \\ & (0.90) \quad (0.032) \\ n = 1641 \quad R^2 = 0.508 \quad F = 186.79 \end{aligned}$$

This is the same regression as presented in footnote. The sample of observations is the same. The only difference is the inclusion of the log³³ (hours worked) on the independent side of the equation. The coefficient for this new variable is now 0.769 i.e. nearer one thus proving the point made above. This has happened probably because we are now trying to constrain the hours worked income to pass through the origin by including some observations with zero income and no hours worked. If we had some intermediate observation or part time, workers the coefficient may move even closer to one.

³¹ These are mainly dairy worker.

³² In Thailand Blaug had an interesting result for occupations. In this analysis amongst the highest paying occupations was the military. The political and military climate is most underdeveloped countries seem to support this thesis. However for lack of any data we cannot firmly ascertain this.

As expected the coefficient for both females and singles are negative. For the former the result is probably indicative of discrimination while for the latter of self-selection process. For females in fact a further testable hypothesis would be that not only does the market discriminate against them in terms of salaries and wages but also in terms of the number of employment opportunities open to them.

An additional unpaid family helper adds about 23% to an individuals income. This figure however may be taken as an indication of the helpers productivity and therefore of their contribution to family income.

A comparison of regression 4 to regression 1 show that most of the explained variance of earnings is accounted for by the basic variables, age, sex, education and marital status³³. The addition of employment status^{immigration} and occupation to these variables raises the R^2 from 0.337 to 0.432 i.e., an addition of only about 0.095 to the explanation. Regression 3 using only the human capital variables age and education yields an R^2 of 0.254. About 59% of the explained variance is therefore due to these variables³⁴. Clearly an important conclusion of the analysis is that a large portion of the inequality in earning is attributable to differing levels of human capital that individuals acquire.

³³In fact the most important variables are age, education and sex in that order. The effect of marital status is not very significant as indicated by the beta coefficients.

³⁴The correct specification however has not been used for the human capital variables. The other equations are all in the wage rate form i.e. controlling for hours worked whereas regression 3 is not. Given the wage rate specification. We would have had an R^2 of at least 0.28 increasing the importance of the human capital variables.

Table 3

	Regression 1	Rank	Regression 4	Rank	Regression 7	Rank
	log monthly income		log monthly income		log monthly wage rate	
Age	1.176	1	1.304	1	1.198	
Age ²	-1.123	2	-1.129	2	-1.109	
Sex (Female)	-0.213	5	-0.222	5	-0.415	
Marital Status(Single)	-0.084	12	-0.109	7	-0.062	
Education 1 Primary	0.088	10	0.082	8	0.080	
Education 2 Secondary	0.258	4	0.278	4	0.281	
Education 3 Higher	0.271	3	0.310	3	0.293	
Tech. Education 2	0.036	21	0.015	9	0.126	
Migration 1 1-3 Years	0.059	17			0.034	
Migration 2 1 if 4-6 Yrs	0.070	14			0.072	
Migration 3 1 if 7-15 "	0.069	15			0.044	
Employment Status 1	0.017	23			0.002	
Employment Status 2	0.173	6			0.141	
Employment Status 3	-0.038	20			-0.039	
Occupation 1 professional & Tech.	0.078	13			0.124	
Occupation 2 1 if 2 Adm & Management	0.078	11			0.082	
Occupation 3 1 if 3 Clerical & related	0.039	19			0.076	
Occn 4 Sales workers	0.064	16			0.011	
Occn 5 Service workers	0.041	18			0.038	
Occn 6 Agricultural	0.021	22			0.024	
Hours worked	0.094	9	0.132	6	0.126	
Years on Job	0.157	7			0.126	
No of unpaid family helpers.	0.129	8			0.144	

As is well known that size of the least squares coefficient in the classical linear model, is not a reliable measure of the relative importance of an independent variable in determining the variation in the regressand. The size of these coefficients can easily be varied by changing the units of measurement. Of the three objective measures of the size of a coefficient we have chosen the beta-coefficients. Table 3 presents the beta-coefficients for 3 of our regressions. The coefficients have been ranked according to size. The ranking reveals age and age² as being the foremost in affecting income distribution followed closely by the two education categories, secondary and higher. Thereafter the rankings show sex, the richer self-employed, unpaid family helpers, years on the job, hours worked, primary education, administrative occupation, marital status migration 2 and the remaining occupation and employment status variables. Thus again we find that age and education the two human capital variables are the most important in affecting the distribution of income.

SOME ADDITIONAL FINDINGS

The regression results of the last section revealed that the bulk of the self-employed i.e. those without employees (SE1) earn no more than regular employees (RE). The self-employed with employees (SE 2) make about 58% more and the casual (CE) make 20% less than the RE. In the rankings the beta-coefficients the SE1 and the CE were among the lowest. There is no indication however, of how well the model used explains the variation in earnings within these categories. The importance of the human capital variables in the previous results may have been due to the larger proportion of the RE in our sample (60%). For the SE and the CE, most probably the structural variables or some other unidentified forces may explain the

differential. The reasoning of the theory of human capital in fact, implies that the acquisition of capital in this form is more likely to benefit the RE than the others³⁵.

There is therefore a need to examine in detail, the results of the previous section especially with regard to employment status. For this reason four subsamples were selected, one for each employment status³⁶. For each subsample then, two regressions were run. The results are presented in table 4 and the salient features of these are discussed below:-

- (i) Note first that in the CE category there are very few observations, too few in fact for a meaningful analysis - The high R^2 's are not as much an indication of a good fit as of the lack of a reasonable number of observations. At best the results for this category are highly unreliable and will not be stressed.
- (ii) Not surprisingly, the model fits the RE the best and the SE 1 the least. In the earlier regressions for all the earners, therefore the greater part of the explanation was due to the presence of the RE. The R^2 of 0.361 for the SE however, is better than expected. The model used thus explains well the variation in earnings within this category.
- (iii) In table 2 most of the explained variation in incomes is due to the four variables, age, sex, education and marital status. These variables now account for about 9.4 %, 82.3% and 44.4% of this explained variation within the RE SE 1 and the SE 2 respectively. As before in the ranking of the beta-coefficients in each of the employment status categories the age and education variables

³⁵ Remember investment in human capital raises an individuals earnings by raising his productivity. Employers will be willing to pay for higher productivity but customers at a shop may not. The possibility of the person using his human capital to improve his business however always remains. In this case the human capital by x hypothesis would hold amongst the SE too.

³⁶ Alternatively, we could have used interaction terms but this procedure would have multiplied to an uncontrollable limit the number of variables.

-:31:-

Table 4

Table 4	Regular Employees				Self employed without emp-				Self employed with employees				Casual			
	Mean (Std dev)	Log monthly income	Log monthly income	Log Monthly income	Mean (Std dev)	Log Monthly income	Log Monthly income	Log Monthly income	Mean Log (Std dev)	Log Monthly income	Log Monthly income	Log Monthly income	Mean n (Std dev)	Log Monthly income	Log Monthly income	
	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	11.	12.	12.	
X ₁ Age	34.986 (12.46)	0.063 (0.008)	0.070 (0.008)	40.74 (14.836)	0.046 (0.010)	0.049 (0.010)	46.1 (9.605)	0.130 (0.078)	0.076 (0.073)	30.130 (12.440)	0.145 (0.066)	0.18 (0.08)				
X ₂ Age		-0.0007 (0.00009)	-0.0008 (0.00009)	- (0.00009)	-0.0005 (0.0001)	-0.0005 (0.0001)	- (0.0001)	-0.0013 (0.0008)	-0.0006 (0.0008)	- (0.0008)	-0.0016 (0.009)	-0.00 (0.00)				
X ₁₂ Sex (Female)	0.062 (0.241)	-0.459 (0.069)	-0.420 (0.067)	0.057 (0.233)	-0.755 (0.130)	-0.829 (0.131)	- (0.0001)	- (0.0008)	- (0.0008)	0.174 (0.388)	-1.767 (0.451)	-0.97 (0.57)				
X ₁₃ Marital Status (Single)	0.341 (0.414)	-0.086 (0.043)	-0.111 (0.044)	0.242 (0.429)	-0.226 (0.079)	-0.257 (0.081)	0.033 (0.180)	-0.069 (0.434)	0.070 (0.427)	0.522 (0.511)	-0.076 (0.412)	0.42 (0.53)				
X ₁₄ Education (Primary)	0.293 (0.455)	0.171 (0.044)	0.219 (0.044)	0.326 (0.469)	0.109 (0.064)	0.084 (0.065)	0.311 (0.467)	0.219 (0.185)	-0.042 (0.172)	0.478 (0.511)	-0.631 (0.224)	-0.18 (0.30)				
X ₁₅ Education (Secondary)	0.388 (0.480)	0.458 (0.049)	0.563 (0.043)	0.129 (0.336)	0.262 (0.092)	0.186 (0.090)	0.197 (0.401)	0.700 (0.259)	0.214 (0.210)	0.087 (0.288)	-0.404 (0.560)	0.41 (0.57)				
X ₁₆ Education (Higher)	0.083 (0.276)	0.840 (0.074)	1.041 (0.065)	0.018 (0.161)	0.746 (0.215)	0.530 (0.214)	0.082 (0.473)	0.582 (0.434)	0.539 (0.286)	- (0.0001)	- (0.0001)	- (0.0001)				
X ₁₈ Technical Education	0.260 (0.439)	0.024 (0.038)	0.018 (0.036)	0.205 (0.404)	0.130 (0.080)	0.093 (0.071)	0.246 (0.434)	0.234 (0.207)	-0.020 (0.173)	0.304 (0.470)	0.187 (0.447)	0.23 (0.42)				
X ₁₉ Migration 1 1-3 Years	0.061 (0.239)	0.177 (0.067)	- (0.0001)	0.018 (0.135)	0.110 (0.209)	- (0.0001)	- (0.0001)	- (0.0001)	- (0.0001)	- (0.0001)	- (0.0001)	- (0.0001)				
X ₂₀ Migration 2 4-6 Years	0.056 (0.229)	0.283 (0.068)	- (0.0001)	0.027 (0.161)	0.079 (0.174)	- (0.0001)	0.049 (0.218)	0.551 (0.343)	- (0.0001)	0.130 (0.344)	0.542 (0.408)	- (0.0001)				
X ₂₁ Migration 3 7-15 "	0.159 (0.366)	0.099 (0.043)	- (0.0001)	0.182 (0.387)	0.084 (0.074)	- (0.0001)	0.098 (0.300)	-0.033 (0.290)	- (0.0001)	0.217 (0.422)	0.175 (0.459)	- (0.0001)				
X ₂₂ Migration 4 16-30 "																
X ₂₇ Occupation 1 Profession & technical	0.145 (0.352)	0.196 (0.059)		0.041 (0.198)	-0.345 (0.156)		0.066 (0.250)	0.077 (0.459)	- (0.0001)	0.087 (0.288)	2.370 (0.578)	- (0.0001)				
X ₂₈ Occupation 2 administrative and managerial	0.020 (0.139)	0.475 (0.118)		0.012 (0.110)	0.498 (0.256)		0.066 (0.250)	-0.465 (0.452)	- (0.0001)	0.0 (0.0001)	- (0.0001)	- (0.0001)				

X ₂₉	Occupation 3 Clarial	0.277 (0.448)	0.026 (0.048)	0.002 (0.045)	-0.046 (0.612)	0.016 (0.128)	0.191 (0.580)	0.0	
X ₃₀	Occupation 4 Sales work	0.074 (0.262)	-0.036 (0.064)	0.480 (0.500)	(0.131) (0.075)	0.344 (0.479)	-0.069 (0.198)	0.087 (0.288)	0.138 (0.397)
X ₃₁	Occupation 5 Service work	0.142 (0.349)	-0.090 (0.052)	0.041 (0.198)	-0.125 (0.126)	0.155 (0.321)	-0.332 (0.257)	0.087 (0.288)	-0.277 (0.608)
X ₃₂	Occupation 6 agricul- tural work	0.007 (0.086)	-0.111 (0.180)	0.105 (0.306)	0.122 (0.150)	0.016 (0.128)	1.206 (0.539)	0.0	-
log									
X ₃₃	Hours worked	5.219 (0.349)	0.070 (0.046)	0.0415 (0.047)	0.477 (0.100)	0.564 (0.101)	0.350 (0.223)	0.031 (0.414)	5.265 (0.297)
X ₃₄	Years on the job	9.342 (8.472)	0.011 (0.002)	13.492 (12.320)	0.006 (0.003)	15.607 (10.190)	0.021 (0.009)	8.00 (6.941)	0.218 (0.021)
X ₃₅	No of unpaid family			0.232 (0.557)	0.261 (0.051)	0.377 (0.489)	0.051 (0.114)	0.0	
K		19	9	20	9	18	8	13	8
R ²	Intercept	3.852	3.94	2.172	1.839	1.242	4.329	2.771	-0.731
n		0.501	0.453	0.361	0.288	0.442	0.193	0.857	0.541
F		809	809	488	488	61	61	23	23
$\frac{F}{R}$		41.76	73.54	13.20	21.53	1.85	1.56	4.13	2.06
		0.490	0.440	0.335	0.276	0.221	0.086	0.685	0.327
	Mean Income	417.216		440.77		982.95		270.22	
	Std. dev. Long Income	0.601		0.739		0.576		0.777	

(Standard errors in Parenthesis)

dominate the others³⁷. The conclusion therefore would be that the human capital variables are important in explaining the income differential in all the employment status categories. A qualification however may be added, that the applicability and the effectiveness of the human capital model is greater among the RE than the others.

The age and education coefficients turn out as expected for each of the subsamples. The age-earnings profiles are all concave and peak at 45 for the RE, SE1 and CE and at 50 for the SE 2. Not surprisingly, the coefficients for education are the largest for the RE.

- (v) Most of the SE 1 and SE 2 are sales workers, while the RE are mainly clerical workers. In view of the description of Rawalpindi given above, this result is hardly surprising. The most paying occupation however is still the administrative and managerial.
- (vi) The mean income is the highest for the SE 2m but for these individuals we know a part of their income is a return on capital. Between the SE1 and the RE, the former have a slightly higher mean income, (a difference of Rs. 25.6). It was because of this difference that in the full sample regression the coefficient for the SE1 was small, positive and insignificant.
- (vii) Inequality as given by the standard deviation of the logarithm of incomes is the least amongst the SE2 and the most amongst the CE. Income amongst the SE1 however is more unequally distributed than among the RE. For the SE1 and the CE the inequality index is greater than the equivalent index for the full sample.

³⁷For the self-employment the sex and unpaid family helpers have a larger beta-coefficient than some of the education dummies. For regular employees however the human capital variable are much stronger. As before marital status is the least important.

(viii) As mentioned earlier the SE 1 and the CE constitute the informal sector while the RE the formal. Given this our results can serve to provide a tentative verification of the dual economy hypothesis. According to this theory there are barriers to entry into the formal sector for those in the informal sector. The latter would like to move into the former sector for incentives like higher earnings, shorter working hours, job security etc. According to our results however they may not get higher earning in the formal sector. But as expected incomes are more equally distributed in the formal sector than in the informal.

(ix) One of the derived variables in the data was the formal/informal. The formal sector was defined as, the government and municipality employees, the professionals, employees of large scale manufacturing and other firms (more than 20 employees). A cross tabulation of this variable with the employment status is shown below:-

	<u>ESI</u>	<u>ES2</u>	<u>RE</u>	<u>CE</u>
Formal 27%		13.3%	70.7%	9.1%
Informal 97.3%		86.7%	29.3%	90.9%

There are 29.3% of RE who should be classified in the informal sector. To the extent that these people are migrants and low-income - earners the above results will be biased.

CONCLUSIONS

It is quite possible when using ordinary least squares in the classical linear model to get spurious results. In particular biased results are likely to result if (a) the regression errors are not randomly distributed with zero mean and constant variance and (b) if these errors are not uncorrelated with each other and with the dependent variable^{z₀}. These doubts however were easily dispelled when the residuals in standardised form, for our main regression (regression 1) were plotted against the estimated value of income, against standardised. A symmetrical spherical pattern emerged indicating that there was no departure from the assumptions of homoskedasticity and independence.

When using such a large number of variables especially as most of them are in the binary form, there is a likelihood of multicollinearity affecting the results. The presence of this problem would bias both the size of the regression coefficients and the value of their standard errors. No special tests for collinearity were conducted but several checks were made to detect its presence. First an inspection of the correlation matrix for the independent variables showed that there was no cause for alarm. Second each of the independent variables was in turn regressed on the other independent variables to test for linear dependence. The R^2 was mostly below 0.30 and never more than 0.55. Third the size of the regression coefficients was not changed much by the addition or deletion of variables from regressions. The presence of multicollinearity on the other hand would have caused violent changes in the size of the coefficients with changes in the number of variables. We may conclude therefore multicollinearity does not

³⁸ On the other hand our results may be suffering from a simultaneous equation bias. The occupation, employment status, education etc. may not all be predetermined variables but some may be choice variables. We would therefore have a simultaneous equation model on our hands of which we have estimated one under-identified equation. In this case the wrong estimation technique ordinary least squares was used. It is not however clear what the correct simultaneous equation model is. A lot would depend on what assumptions one makes, which variables are the endogenous ones etc. For our purposes however the single equation model is all right for we take the characteristics of the individual as given and ask the question how these effects his income.

affect our results.

From our analysis we may therefore conclude that the human capital variables explain a large part of the income differential³⁹. Age and education both affect income as expected. Given greater equality in education therefore it would be reasonable to expect more closely grouped together age-earnings profiles and therefore reduced inequality. Interestingly enough the human capital variables explained the larger part of the variation in earnings. Returns to education followed the same pattern across sectors; increasing returns to increasing levels of education.

For the dual market hypothesis our results show that individuals in the informal sector earn about as much or may be slightly more than those in the formal. Income however is more unequally distributed in the former sector than in the latter. The barriers to entry if any however we cannot identify in this analysis. But it seems as if there is no financial incentive for moving from the informal sector to the formal. Two incentives for such a move can however be identified from our analysis; (a) shorter working hours and hence opportunity for secondary employment, (b) a more equally distributed guaranteed monthly earnings. There are however a number of other incentives which one can hypothesise like freedom from the vicissitudes of the market, old age pension, paid holidays and other such employment benefits.

The results for migration were interesting enough to deserve a mention here. The pattern of earnings for migrants was concave to the relatively settled earned the most followed by the recent arrivals and

³⁹Generalisations beyond the city however must be made bearing in mind the socio-economic description of the city given in section 3. The lack of large scale manufacturing industry and the unusually large presence of the government sector in the form of Islamabad the capital probably biases the results in favour of the human capital variables.

"the settled". However migrants every where earned more than the local residents. Contrary to the dual economy theory the majority of the migrants were found in regular employment. In fact 85% of the recent migrants are in regular employment. These individuals are not therefore as theory predicts using the informal sector as a point of entry.

To conclude therefore, our main result is the applicability and importance of the human capital variables in explaining the income differential. Both work experience and higher levels of education seem to command a premium in the market. But do the educated earn more because they are more productive? or more able? or because they happen to belong to influential families? Unfortunately none of these questions can be assumed with any degree of accuracy by our analysis. Data on family background was not available, and so the effects of parental education, income etc. on earnings and education could not be observed. Ability as is obvious would be an exceedingly difficult variable to measure. No assessments of the filtering process of education, i.e. the sorting out of the more able from the lesser can therefore be made.

Bhatti/*
28977/

REFERENCES

1. Anand, S. The size distribution of income in Malaya.
Oxford (unpublished)
2. Becker, Gary S. Human Capital New York : Columbia University
Press . 1974.
3. Blaug, M. An Economic Analysis of Earnings in Thailand
in Economic Development and Cultural Change Vol. 23
Number 1 Oct. 1974.
4. Blaug, M. Rate of Return to Investment in Education in
Thailand Published by National Education
Council Thailand.
5. Chiswick, Barry. Income inequality New York : Columbia University
R. Press 1974.
6. Goldberger, A.S. Econometric Theory Thon Willey and Sons. Inc.
New York.
7. Hamdani, K. Education and the Income differential: An
Estimation for Rawalpindi City.
Pakistan Development Review 1977.
8. Masumdar, D. The urban informal sector.
World Bank Staff Working Paper No. 211.
9. Sen, A. K. On Income Inequality Oxford
Clarendon Press 1973.

This work is licensed under a
Creative Commons
Attribution – NonCommercial - NoDerivs 3.0 Licence.

To view a copy of the licence please see:
<http://creativecommons.org/licenses/by-nc-nd/3.0/>